

CORRECTING SEGMENTATION ERRORS IN OCR

ABSTRACT

A method for encoding characters includes identifying one or more sequences of the character codes
5 that are likely to be generated due a segmentation error
in application of a pattern recognition process, and
associating a respective extension character code with
each of the sequences. The area of an image containing
characters is divided into segments, such that each
10 segment contains approximately one character. The
pattern recognition process is applied to each of the
segments in order to generate an input string of
character codes. At least one of the identified
sequences of the character codes in the input string is
15 replaced with the respective extension character code so
as to generate a modified string. The output string is
determined by comparing the modified string to a
directory of known strings.